

Serial No.: 09/729,240  
Docket No.: 0026-0001

REMARKS

Applicants appreciate the Examiner's indication that claims 4, 9, 16, 23, 28, and 33 are directed to allowable subject matter.

Additionally, in the Office Action of May 8, 2003, the Examiner rejected claims 1-3, 5-8, 10-15, 17-22, 24-27, 29-32, 34, and 36-41 under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 5,778,361 to Nanjo et al. ("Nanjo) in view of U.S. Patent No. 6,513,031 to Fries et al. ("Fries") and U.S. Patent No. 6,088,692 to Driscoll ("Driscoll"). Further, the Examiner rejected claim 35 under 35 U.S.C. § 103(a) as being unpatentable over Nanjo, Fries, Driscoll, and further in view of U.S. Patent No. 6,134,554 to Freimann et al. ("Freimann").

For the following reasons, Applicants respectfully traverse the rejections based on Nanjo, Fries, Driscoll, and Freimann.

Applicants submit that claim 1, for example, is not disclosed or suggested by the combination of Nanjo, Fries, and Driscoll. Claim 1 is directed to a method of identifying semantic units within a search query. The term "semantic unit," as used in the pending application, refers to multiple terms that are considered to function as a "compound" that forms a single semantically meaningful unit.

(Spec., page 2). Semantic units are identified in claim 1 through a method that includes identifying documents relating to a search query by matching individual search terms in the query to an index of a corpus and generating multiword substrings of the query in which each of the substrings includes at least two words. For each of the generated substrings, a value is calculated that corresponds to a comparison between one or more of the identified documents

Serial No.: 09/729,240  
Docket No.: 0026-0001

and the generated substring. Semantic units are selected from the generated multiword substrings based on the calculated values.

The Examiner relies on Nanjo to disclose identifying units within a search query. (See Office Action, numbered paragraph 4). The Examiner concedes, however, that Nanjo does not disclose identifying semantic units and multiword substrings from a query. (*Id.*). For these features of claim 1, the Examiner relies on Fries and Driscoll. In particular, the Examiner appears to rely on Fries for disclosing a multiword substring and on Driscoll for disclosing the identification of semantic units.

Regarding Fries, the Examiner points to column 10, lines 13-21 and column 12, lines 51-64 as disclosing multiword substrings that include at least two words. (Office Action, page 3). These sections of Fries disclose the operation of a natural language processor (NLP) component 227 that performs a natural language parse of a search query. Regarding the NLP component, Fries discloses:

NLP component 227 can also group together multiple words that represent a single conceptual item. For instance, NLP is able to identify the constituent parts of a date as belonging to a single date construct. To identify these "multi-word entries", NLP component 227 utilizes "factoids" and "captoids". Factoids are rules that identify multi-word entries on the basis of known facts. For example, NLP component 227 identifies "New Jersey" as a single multi-word entry because of the fact that New Jersey is a state. Captoids are rules that identify multi-word entries on the basis of the capitalization of terms in the query. For instance, if "Jack's Seafood Restaurant" is found in a search query, NLP component 227 will identify it as a multi-word entry on the basis of its capitalization.

Serial No.: 09/729,240  
Docket No.: 0026-0001

(Fries, column 10, lines 13-26). Thus, it appears to Applicants that the multiple words that represent a single conceptual item, as discussed by Fries, are in fact a semantic unit. These "multi-word entries" in Fries are explicitly stated as being identified based on "factoids" and "captoids." These factoids and captoids appear to be determined based on a predefined set of rules.

Claim 1 is directed to a method of identifying semantic units within a search query. In contrast to Fries, however, the semantic units of claim 1 are not identified based on factoids or captoids. Instead, the semantic units are identified by identifying documents relating to a query by comparing search terms in the query to an index of a corpus; generating a plurality of multiword substrings from the query in which each of the substrings includes at least two words; calculating, for each of the generated substrings, a value that corresponds to a comparison between one or more of the identified documents and the generated substring; and selecting semantic units from the generated multiword substrings based on the calculated values. Although Fries may disclose identifying semantic units as "multi-word entries," the rule-based factoid and captoid techniques of Fries for identifying these entries in no way discloses or suggests the features of claim 1.

Applicants further submit that Driscoll does not cure the deficiencies of Nanjo and Fries. As discussed in the previous response, Driscoll is directed to a natural language search system and method for searching and ranking relevant documents from a database. A search query in Driscoll is used to generate a group of documents. (Abstract). Each word in the search query and the

Serial No.: 09/729,240  
Docket No.: 0026-0001

documents is assigned a weighted value. (*Id.*). The weighted values are then used to generate a similarity value by which the documents may be ranked. (*Id.*).

The search techniques in Driscoll make use of "semantics" (Driscoll, col. 3, lines 36-37) or "the concept of semantics" (Driscoll, col. 3, lines 19-23) to improve search results. Exemplary semantic units in Driscoll are shown in Figs. 8 and 9A-9E. As shown, the semantic units of Driscoll are all single words that are looked up in a thesaurus. (Driscoll, col. 6, lines 39-47). The thesaurus associates each word with one or more categories. The word categories are used by Driscoll when determining the relevance or similarity of a document to a query. (Driscoll, col. 5, lines 43-48).

Thus, Applicants submit that to the extent Driscoll identifies semantic units, these units are identified in a manner significantly different from the present invention. Where Fries uses predefined rules that define factoids and captoids, Driscoll uses a thesaurus to identify semantic units. In contrast to either Driscoll or Fries, however, the present invention, as recited in claim 1, identifies semantic units in a manner significantly different from either Driscoll or Fries. That is, as recited in claim 1, semantic units are identified in a search query by identifying documents related to the search query, generating multiword substrings from the query, calculating values that correspond to the comparisons between the identified documents and the generated substrings, and selecting semantic units from the generated multiword substrings based on the calculated values.

At page 4 of the Office Action, the Examiner contends that Driscoll discloses:

Serial No.: 09/729,240  
Docket No.: 0026-0001

calculating, for each of the generated substrings, a value that corresponds to a comparison between one or more of the identified documents and the generated substrings (see column 6, lines 53-64); and selecting semantic units from the generated multiword substrings based on the calculated values (see column 4, lines 27-32).

(Office Action, numbered paragraph 4, page 4, lines 9-12). Applicants disagree with this characterization of Driscoll. As mentioned above, Driscoll selects semantic units based on a thesaurus lookup. Driscoll does not, however, select semantic units from generated multiword substrings based on the calculated values recited in claim 1. The section of Driscoll cited by the Examiner as allegedly disclosing this feature of claim 1 states:

FIG. 11 is a list of the eight documents in the example of FIG. 3 ranked in order of their relevance or similarity (SIM value) to the words used in the original query of FIG. 3 and shown in FIG. 5; both the DocId and the SIM value are shown as a pair in this list. This list is a sorted Relevancy List created at Step 900 in FIG. 1.

(Driscoll, col. 4, lines 27-32). This section of Driscoll completely fails to mention semantic units, and does not appear to relate to selecting semantic units from a generated multiword substring, as recited in claim 1. If the Examiner maintains the current rejection of claim 1, Applicants request additional clarification on how the Examiner is interpreting this section of Driscoll in relation to claim 1.

A proper rejection under 35 U.S.C. § 103(a) requires that the examiner establish a *prima facie* case of obviousness. To establish a *prima facie* case of obviousness, three basic criteria must be met. First, there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to modify the reference or to

Serial No.: 09/729,240  
Docket No.: 0026-0001

combine reference teachings. Second, there must be a reasonable expectation of success. Finally, the prior art reference (or references when combined) must teach or suggest all the claim limitations. The teaching or suggestion to make the claimed combination and the reasonable expectation of success must both be found in the prior art, not in Applicants' disclosure. In re Vaeck, 947 F.2d 488, 20 USPQ2d 1438 (Fed. Cir. 1991). (See also, M.P.E.P. 2143).

Instead of establishing a proper *a prima facie* case of obviousness, the Examiner appears to be arbitrarily picking and choosing various features from Nanjo, Fries, and Driscoll based on Applicants' disclosure. Applicants remind the Examiner that the teaching or suggestion to make the claimed combination and the reasonable expectation of success must both be found in the prior art and not in Applicants' disclosure. If anything, because Driscoll and Fries each disclose different methods of identifying semantic units, these patents tend to teach away from the present invention, as one of ordinary skill in the art reading Driscoll or Fries would tend to use the techniques disclosed by these patents to identify semantic units and not that of the present invention.

For at least these reasons, Applicants submit that the rejection of claim 1 under 35 U.S.C. § 103(a) is improper and should be withdrawn. At least by virtue of their dependency on claim 1, the rejection of claims 2, 3, 5, and 37 should also be withdrawn. In addition, these claims recite additional features neither disclosed nor suggested by the combination of Nanjo, Fries, and Driscoll.

For example, claim 3 further defines the method of claim 1, and recites that the selection of semantic units further includes "selecting semantic units from

Serial No.: 09/729,240  
Docket No.: 0026-0001

the generated substrings that have calculated values above a predetermined threshold." The Examiner points to column 20, lines 41-50, of Nanjo as disclosing this feature. (Office Action, page 5). This section of Nanjo corresponds to a feature of claim 1 of Nanjo that refers to "step indexing the symbols in the preliminary index term to create a plurality of index terms of a length equal to or less than a predetermined step size."

Applicants submit that the predetermined step size recited in this claim of Nanjo is not equivalent to, and does not disclose or suggest selecting semantic units based on calculated values above a predetermined threshold. Nanjo merely creates a plurality of index terms by stepping through a preliminary index term using a predetermined step size. This does not disclose or suggest, however, comparing calculated values for substrings to the predetermined threshold recited in claim 3. Accordingly, for this reason, as well as the dependency of claim 3 to claim 1, the rejection of claim 3 should be withdrawn.

Claim 5 further defines the method of claim 1, and recites that:

the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant ones of the identified documents are assigned higher calculated values than substrings that occur in less relevant ones of the documents.

The Examiner points to column 6, lines 1-64, of Driscoll as disclosing this feature of claim 5. (Office Action, page 6). This section of Driscoll describes the calculation of the "SIM value," which measures the relevance of a document to a query. Calculating the relevance of a document to a query, as disclosed by Driscoll, does not disclose or suggest the calculated values recited in claim 5, in

Serial No.: 09/729,240  
Docket No.: 0026-0001

which substrings that occur in more relevant ones of the identified documents are assigned higher calculated values than substrings that occur in less relevant ones of the documents. For at least this additional reason, the rejection of claim 5 should be withdrawn.

Independent claim 6 is directed to a method for locating documents in response to a search query. Claim 6 recites a number of features similar to those recited in claim 1, including "generating a plurality of multiword substrings of the query" and "selecting semantic units from the generated multiword substrings based on the calculated values." For reasons similar to those given above, Applicants submit that Nanjo, Fries, and Driscoll, either taken alone or in combination, do not disclose or suggest these features of claim 6.

Accordingly, for at least these reasons, Applicants submit that claim 6 is not disclosed or suggested by Nanjo, Fries, and Driscoll, either alone or in combination. The rejection of this claim should thus be withdrawn. At least by virtue of their dependency from claim 6, the rejection of claims 7, 8, 10, and 38 should also be withdrawn.

Additionally, dependent claims 8 and 10 recite features similar to those recited in dependent claims 3 and 5, respectively. Accordingly, for reasons similar to those given above regarding claims 3 and 5, the rejection of claims 8 and 10 should be withdrawn.

Independent claim 11 is directed to a system and includes features similar to those recited in claim 1. Thus, for reasons similar to those given with respect to claim 1, the rejection of claim 11 should also be withdrawn.

Serial No.: 09/729,240  
Docket No.: 0026-0001

The rejection of claims 12-15, 17, and 39, which depend from claim 1, should also be withdrawn, at least by virtue of their dependency.

Additionally, dependent claim 15 recites features similar to those recited in dependent claim 3. Accordingly, the rejection of claim 15 should also be withdrawn for reasons similar to those given above regarding claim 3:

Independent claim 18 recites a number of features, including a ranking component configured to return a list of documents ordered by relevance in response to a search query and a semantic unit component configured to locate semantic units, having a plurality of words, in search queries entered by a user based on a predetermined number of most relevant documents in the list of documents returned by the ranking component. As previously discussed, neither Nanjo, Fries, nor Driscoll discloses or suggests locating semantic units in search queries based on a predetermined number of most relevant documents in the list of documents returned by the ranking component. Accordingly, the rejection of claim 18 is improper and should be withdrawn.

The rejection of claims 19-22 and 24, at least by virtue of their dependency from claim 18, either directly or indirectly, should also be withdrawn.

Additionally, dependent claim 22 recites features similar to those recited in dependent claim 3. Accordingly, the rejection of claim 22 should be withdrawn for reasons similar to those given above regarding claim 3.

Independent claim 25 recites features similar to those recited in claim 1. Independent claims 30 and 36 recite features similar to those recited in claim 6.

Serial No.: 09/729,240  
Docket No.: 0026-0001

Thus, for reasons similar to those given above regarding claims 1 and 6, the rejection of these claims should also be withdrawn. The rejection of claims 26, 27, 29, 31, 32, 34, 40, and 41 at least by virtue of their dependency from one of claims 25 or 30, should also be withdrawn.

In rejecting dependent claim 35, the Examiner relies on Freiman, in addition to Nanjo, Driscoll, and Fries, for the disclosure of a computer-readable medium. Applicants have reviewed Freiman, and submit that the disclosure of Freiman does not cure the above-noted deficiencies in the disclosures of Nanjo, Driscoll, and Fries. Accordingly, the rejection of claim 35 should also be withdrawn.

In view of the foregoing remarks, Applicants submit that the claimed invention is not rendered obvious in view of the prior art references cited against this application. Applicants therefore request the Examiner's reconsideration and reexamination of the application, and the timely allowance of the pending claims. If the Examiner believes that the application is still not in condition for allowance, Applicants invite the Examiner to call the undersigned at the below-listed number to discuss any further issues that the Examiner may have.

**OFFICIAL**Serial No.: 09/729,240  
Docket No.: 0026-0001

To the extent necessary, a petition for an extension of time under 37 C.F.R. § 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account 50-1070 and please credit any excess fees to such deposit account.

**RECEIVED**

AUG 11 2003

Technology Center 2100

Respectfully submitted,

HARRITY &amp; SNYDER, L.L.P.

By: Brian E. Ledell  
Reg. No. 42,784

Date: August 8, 2003

11240 Waples Mill Road  
Suite 300  
Fairfax, Virginia 22030  
Telephone: 571-432-0800  
Facsimile: 571-432-0808